

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



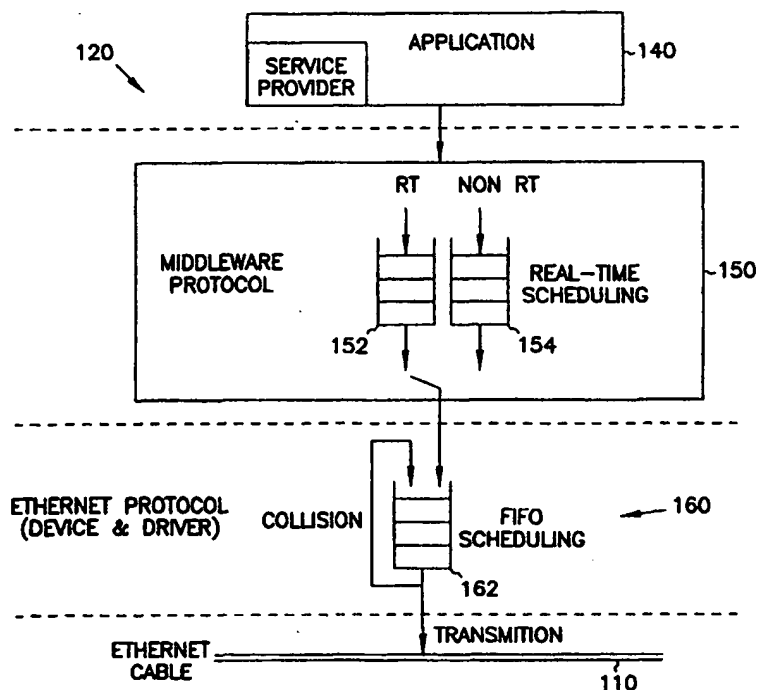
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : H04L 12/64, 12/413		A1	(11) International Publication Number: WO 00/03521
			(43) International Publication Date: 20 January 2000 (20.01.00)
(21) International Application Number: PCT/US99/14083		(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 22 June 1999 (22.06.99)			
(30) Priority Data: 09/113,732 10 July 1998 (10.07.98) US		Published With international search report.	
(71) Applicant: HONEYWELL INC. [US/US]; Honeywell Plaza, Minneapolis, MN 55408 (US).			
(72) Inventors: HUANG, Jiandong; 4365 Juneau Lane, Plymouth, MN 55446 (US). CHEN, Donghui; 425 13th Avenue S.E., Minneapolis, MN (US).			
(74) Agent: MACKINNON, Ian, D.; Honeywell Inc., Honeywell Plaza - MN12-8251, P.O. Box 524, Minneapolis, MN 55440-0524 (US).			

(54) Title: MIDDLEWARE-BASED REAL-TIME COMMUNICATION SYSTEM

(57) Abstract

A middleware approach to implementation of real-time Ethernet that provides deterministic, i.e. predictable, communication services over the conventional Ethernet network is described. The invention resides above the network interface device and the device driver, yet below the system transport services and/or user applications. The middleware schedules and controls admission of data packets onto the network and guarantees the real-time constraints of the data packets once they are admitted. The invention prohibits collision of data streams during transmission of real-time data, yet may allow collisions during transmission of soft- or non-real-time data for improved utilization of bandwidth. The invention may further optimize bandwidth utilization by incorporating a quality of service definition into the scheduling determination.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

MIDDLEWARE-BASED REAL-TIME COMMUNICATION SYSTEM

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by any one of the patent disclosure, as it appears in the Patent and Trademark Office patent files or records, but otherwise reserves all copyright rights whatsoever.

FIELD OF THE INVENTION

The invention is related to network data communications, and in particular to a middleware approach to implementation of real-time Ethernet.

BACKGROUND OF THE INVENTION

Computer networks have become widely popular throughout business and industry. They may be used to link multiple computers within one location or across multiple sites.

The network provides a communication channel for the transmission of data, or traffic, from one computer to another. Network uses are boundless and may include simple data or file transfers, remote audio or video, multimedia conferencing, industrial control and more.

Perhaps the most popular network protocol is Ethernet, a local area network specification for high-speed terminal to computer communications or computer to computer file transfers. The Ethernet communication protocol permits and accommodates data transfers across a data communication channel or bus, typically a twisted pair or coaxial cable.

The Ethernet communication protocol was standardized as the IEEE 802.3 standard for communications over a local area network (LAN). This protocol incorporates a 1-persistent, Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol meaning that one or more nodes of a shared network may monitor the channel if they have a data packet to transmit, and transmit that packet immediately upon detecting the channel to be idle.

A "collision" of data packets may occur if two or more nodes begin transmitting simultaneously on the network. Colliding nodes will detect such a collision of data and terminate their transmission, waiting a randomly-determined time period before

attempting transmission again. Under current standards, a failure will be generated after a node makes sixteen unsuccessful attempts to transmit its data packet without collision.

Under lightly-loaded conditions, collisions are infrequent and resolution is rapid. However, heavy loading may lead to indeterminate access time. While some
5 applications may be relatively insensitive to collisions and their resultant delays on data transfer, other applications may be time sensitive such that collisions of data packets are undesirable or even intolerable. Examples of such time-sensitive or real-time applications may include remote video or control of industrial process equipment. The requirement for some applications to circumvent collisions and guarantee successful
10 transmission and reception has led to various improvements to Ethernet.

One Ethernet improvement is a token-based protocol standardized under IEEE 802.4 (Token bus) or 802.5 (Token ring). The primary difference between these two standards is in the network topology each is designed to address. Token bus addresses a network in which the nodes form a logical ring; Token ring addresses a network in
15 which the nodes form a physical ring.

Token-based protocols generate a "token" which is passed to every node along the network. These protocols permit data transmission only when the node is in possession of the token, and each node is given a fixed amount of time to transmit data. This transmission time is further divided into multiple segments or timers relating to
20 different priority levels. These priority levels may be assigned to different data streams depending upon their criticality and time sensitivity. Nodes may only transmit data of a given priority level during its respective timer. Under this approach, real-time data may be assured a fraction of the bandwidth free of collision. However, some of these token-based protocols may allow a given node only its fixed share of bandwidth regardless of
25 whether other nodes make full or even any use of their bandwidth.

Improvements on these token-based protocols have also been proposed. As an example, an academic prototype has been proposed for a software-oriented real-time Ethernet implemented on a UNIX platform utilizing a token-based protocol. (see Chitra Venkatramani, "The Design, Implementation and Evaluation of RETHER: A Real-Time
30 Ethernet Protocol," Ph.D. Dissertation, State University of New York, January 1997) RETHER, however, only provides for non-real time traffic when there is no more real time traffic to be sent by any node. Depending on the type of traffic on the network, this led to low network throughput and utilization due to token passing overhead for non-real time traffic, and did not support hard real time traffic.

Another prior solution is hardware based. Under this approach, data packet collisions are avoided through hardware. These hardware-based solutions may be necessary for certain critical real-time applications such as aviation, to meet stringent performance and reliability requirements. However, such solutions are proprietary and vendor-dependent, making them difficult and expensive to implement. Hardware-based solutions may be incompatible with many existing Ethernet networks, requiring costly and complicated modifications. In addition, although these hardware solutions prevent collisions, they do not offer scheduling of real-time traffic in an entire system. Both solutions also require modification of existing hardware or software.

Accordingly, there exists a need for an efficient deterministic service to prevent collisions of and guarantee real-time traffic over Ethernet that can be implemented on existing Ethernet networks and is compatible with a wide variety of commercial-off-the-shelf (COTS) hardware and applications. Such a solution is needed for process control networks, time sensitive multimedia and Internet applications.

SUMMARY OF THE INVENTION

A middleware approach to implementation of real-time Ethernet provides deterministic, i.e. predictable, communication services for both real time and non-real time traffic over a conventional Ethernet network having a plurality of nodes desiring to transmit packets of data. The middleware comprises computer software residing above a network interface device and the device driver, yet below the system transport services and/or user applications. The invention provides a Middleware Real-Time Ethernet or MRTE which does not require modification to existing hardware that implements Ethernet.

In one embodiment, Ethernet bandwidth is divided into cycles. During each cycle, a first time interval is provided for real time data packet traffic using a deterministic scheduling protocol such as by passing a token, such that no collisions can occur. During a second time interval, the standard carrier sense, multiple access, collision detect Ethernet protocol is used for non-real time traffic. By using these two time intervals, bandwidth is shared between real time and non-real time traffic, ensuring that both will receive desired bandwidth.

In one embodiment, separate queues are used for deterministic scheduling to determine the order of packet queuing and transmission on each node such that (1) real-

time traffic can be guaranteed once admitted for transmission service, (2) non-real-time traffic can be served, and (3) the Ethernet bandwidth utilization can be optimized.

Quality of Service, QoS, enables making on-line tradeoffs between network bandwidth availability and network transmission quality. Examples of QoS include (1) degree of packet collisions when Ethernet is shared by soft- or non-real-time traffics during certain time slots and (2) amount of end-to-end packet transmission latency.

When QoS is used, periodic data, such as video at 30 frames per second may be given a priority or criticality, and a cumulative loss factor, e.g. up to four frames in a row may be discarded. If there is sufficient bandwidth remaining after higher priority tasks or data streams are handled, the video will be accepted to the real time queue with at least five frames per second being sent. If other tasks are deleted or reduced, this frame rate will increase.

Software structuring enables hosting of the real-time Ethernet middleware above the Ethernet network device and the device driver, and below system transport software and/or user applications. A specific example of such a software host is the Microsoft® Network Device Interface Specification (NDIS) with Device Driver Kit (DDK) on Microsoft® NT®-based personal computer platforms. Many other software hosts are available depending upon specific hardware chosen.

A collision avoidance module guarantees that a transmission will not result in traffic collision. The collision avoidance module implements a collision-avoidance protocol that provides the capability for preventing Ethernet traffic from colliding, which is one source of the problem of non-deterministic Ethernet behavior. A specific example of such a protocol is a token-based protocol by which a token circulating among the Ethernet nodes determines which node should transmit packets at any point in time. Other collision-avoidance protocols may be used with the invention such as various implementations of Time-Division Multiple Access (TDMA), a technology using Time-Division Multiplexing (TDM). The protocol or standard provides a mechanism to avoid conflict among data transmission by more than one node at any given time.

In one embodiment, the collision-avoidance protocol is switchable to be enabled or disabled as desired by the deterministic scheduling module. This allows the invention to guarantee no collisions of real-time traffic, yet permit collisions of soft- and non-real-time traffic. Such mixed-mode operation could lead to increased bandwidth utilization depending upon the loading during time periods allocated to soft- and non-

real-time traffic. Lightly loaded CSMA/CD systems can be more efficient than systems operating on a collision avoidance protocol.

While the collision-avoidance protocol is active, the time set for a complete rotation of transmitting nodes is bounded. In the case of a token-based protocol, the token must return within this bounded time, or token rotation time.

For each collision avoidance protocol (token based or TDMA), a deterministic scheduling module uses an algorithm to schedule traffic and guarantee that transmission will be done before a deadline expires.

In a further embodiment of the invention, allocation of bandwidth to an individual bridge or node is increased based on underutilization of bandwidth by other bridges or nodes in the network.

One advantage of the invention is that it remains compliant with the IEEE 802.3 standard. Such compliance allows the invention to be practiced on a multitude of standard Ethernet networks without requiring modification of hardware, thus remaining an open system.

A further advantage of the invention is that it is modular in nature. As such, the invention may be practiced using a variety of collision-avoidance protocols, deterministic scheduling algorithms, and QoS negotiation and adaptation policies and algorithms.

As a software approach, the invention also enables use of any COTS Ethernet cards and drivers for real-time Ethernet. Use of specific vendor Ethernet cards and drivers is transparent to applications, thus making the invention capable of vendor interoperability, system configuration flexibility and low cost to network users.

Brief Description of the Drawings

- Figure 1 is a block diagram of an Ethernet network having multiple nodes incorporating the invention
- Figure 2A is a block diagram of a software architecture of an Ethernet node of Figure 1.
- Figure 2B is a block diagram of the software architecture of Figure 2 providing more detail of certain portions.
- Figure 3 is a diagram depicting the behavior of a node in response to an application request for admission to the network.

- Figure 4 is a diagram depicting the behavior of a node in response to a scheduler interrupt granting admission to the network.
- Figure 5 is a flowchart depicting the behavior of the QoS manager of one embodiment of the invention.
- 5 Figure 6 is a flowchart depicting the behavior of the scheduler of invention.
- Figure 7 is a flowchart depicting the behavior of the MRTE protocol of the invention.
- Figure 8 is a flowchart depicting the interrupt handler of the MRTE protocol of the invention.
- 10 Figure 9 is a flowchart depicting the admission control process of the scheduler of the invention.

DESCRIPTION OF THE EMBODIMENTS

15 In the following detailed description, reference is made to the accompanying drawings which form a part hereof, and in which is shown by way of illustration specific embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized and that structural, logical and electrical changes may be made without departing from the spirit and scope of the invention. The following detailed description is, therefore, not to be taken in a
20 limiting sense, and the scope of the invention is defined by the appended claims. Like numbers in the figures refer to like components, which should be apparent from the context of use.

Figure 1 shows a conceptualized drawing of a simplified Ethernet network incorporating the invention. The Ethernet network 100 comprises an Ethernet channel 110. The Ethernet network 100 further contains two or more nodes 120 which are connected to the Ethernet channel 110 via drops 130 for data receipt and transmission across the Ethernet channel 110. The nodes 120 contain an application layer 140, a middleware real time Ethernet (MRTE) layer 150, and an Ethernet protocol layer 160.
25 The application layer 140 and the MRTE layer 150 are in direct communication and the MRTE layer 150 and the Ethernet protocol layer 160 are in direct communication, but the application layer 140 and the Ethernet protocol layer 160 communicate only through the MRTE layer 150.
30

Figure 2A shows more detail of the layers of a node 120. In particular, MRTE layer 150 logically comprises a pair of queues for data traffic or packets generated in the applications layer 140 for transmission to another node. The first queue comprises a real time queue 152 for queuing information packets that have been accepted for transmission on a real time basis. In other words, packets in this queue are guaranteed to be sent without collision with another packet unless there is a network failure. The real time traffic queue 152 has traffic sorted by criticality. A second queue comprises a non-real time queue for data packets that do not need to arrive at a destination in real time to be of value to the receiving node. The second queue 154 is sorted by first in, first out. The queues may be physically separate or combined with appropriate control software. Both of these queues empty into a standard Ethernet collision queue 162 having a first in, first out scheduling algorithm. Applications in the applications layer 140 may assign data to either of the queues as desired.

A bandwidth partition scheme is implemented such that for a given repetitive cycle of time, MRTE layer 150 implements a deterministic schedule for packets in the real time queue where collisions on the network are avoided for a first time period, and a standard Ethernet protocol during a second time period to allow transmission of non-real time packets obtained from the non-real time queue 154.

In Figure 2B, a node is shown having a more detailed representation of the MRTE layer 150, in particular, showing how the real time queue 152 is managed to provide collision avoidance. In the application layer 140, an application 205 is coupled with a service provider 210. In practice, the application 205 acts as either an ultimate source (sender) or destination (receiver) of data packets. The service provider 210 serves as an interface between the application 205 and the MRTE layer 150.

The MRTE layer 150 is further divided into QoS adaptation services 215 and deterministic scheduling services 220, both of which are implemented in software modules or objects in one embodiment. QoS adaptation services 215 contains a QoS manager 225 and a QoS adaptation algorithm 230. The QoS manager 225 and its associated QoS adaptation algorithm 230 provide QoS-based negotiation and adaptation services such as changing the duration of non-real-time data traffic or suspending low-criticality traffic in order to ensure that sufficient collision free bandwidth is provided for high priority real time traffic as balanced against bandwidth for non-real time traffic.

Deterministic scheduling services 220 contains a collision resolution protocol 235, an MRTE protocol 240, and an MRTE scheduler 245 and its associated

deterministic scheduling algorithm 250 and MRTE repository 255. Deterministic scheduling services 220 further contains software structuring 260 to host or interface the MRTE layer 150 to the Ethernet protocol layer 160. MRTE protocol 240, with the aid of collision resolution protocol 235, provides arbitration to avoid collisions of data packets. There is one protocol per Ethernet configuration serving data packet transmission. MRTE scheduler 245, with the aid of deterministic scheduling algorithm 250, provides scheduling analysis and coordinates distributed scheduling among individual nodes. There is one scheduling algorithm per Ethernet configuration. MRTE scheduler 245 utilizes MRTE repository 255 for the storage of a local image of global scheduling information.

The Ethernet protocol layer 160 contains the Ethernet driver 265 which supports the Ethernet card (not shown) for physical connection and communications to the Ethernet channel 110. Figure 2B depicts bidirectional data flow connecting service provider 210 to MRTE protocol 240 to software structuring 260 to Ethernet driver 265. Figure 2B further depicts control flow connecting the MRTE protocol 240 to the collision resolution protocol 235 and the MRTE scheduler 245. The MRTE scheduler additionally controls flow communication with the applications 205, the QoS manager 225 and the deterministic scheduling algorithm 250. The deterministic scheduling algorithm controls the flow of communication with the QoS manager 225. The QoS manager further controls the flow of communication with the applications 205 and the QoS adaptation algorithm 230.

The various components depicted in Figure 2B can further be described as software objects as shown in Table 1. The individual software objects communicate via application programming interface (API) calls. The calls associated with each object are listed in Table 1.

Table 1
Software Objects and API calls

Object	Responsibility	API
MRTE Service Provider 210	a) MRTE interface to application	Admit() IndicateAdmit()
	b) Buffering for	Send()

	received packets	Receive() IndicateReceive()
QoS Manager 225	a) QoS negotiation b) QoS adaptation	Admit()
MRTE Scheduler 245	a) Deterministic bandwidth scheduling b) Setting up fragmentation parameter	Admit() IndicatePermission()
Deterministic Scheduling Algorithm 250	Provide a scheduling algorithm and conduct schedulability analysis accordingly	Schedule(IN req, IN QoS, OUT frag)
MRTE Protocol 240	a) Arbitration to control i) traffic collision; and ii) scheduling sequence b) Packet fragmentation and transmission via Software Structuring	GetPermission() Update(StreamID, frag, . . .) Send() IndicateReceive() TransferData()
MRTE Repository 255	Local image of global scheduling information	Get() Update()
Software Structuring 260	a) Framework for "plug-n-play" drivers and SPI	Send() TransferData()

Admission control to the network may be initiated in one of two ways.

Admission may be requested by the applications or it may be initiated by a scheduler interrupt. Figures 3 and 4 depict the behavior of the various components in controlling admission.

Figure 3 depicts the behavior in response to an application request. Applications 205 will make an Admit() call on the MRTE service provider 210 to indicate a desire to transmit data across the network. MRTE service provider 210 in turn forwards the Admit() call to the MRTE scheduler 245. MRTE scheduler 245 then invokes a
5 GetPermission() call on the MRTE protocol 240 to indicate that a request for admission is pending.

Figure 4 depicts the behavior in response to a scheduler interrupt. An interrupt is first received by the MRTE protocol 240 in a form dictated by the collision avoidance protocol chosen. For a token-based protocol, the interrupt is received from software
10 structuring 260 and indicates receipt of the token by an individual node. For a TDMA-based protocol, the interrupt is generated by the TDMA timer (not shown) and indicates that the time slot is appropriate for data transmission by an individual node. Upon receipt of the interrupt, MRTE Protocol 240 makes an IndicatePermission() call on the MRTE scheduler 245. MRTE scheduler 245 then makes a Get() call on MRTE
15 repository 255 to get network status information. MRTE scheduler 245 then makes a Schedule() call on the deterministic scheduling algorithm 250. The MRTE repository 255 issues an Update() call to provide the network status to MRTE scheduler 245, which is then forwarded to MRTE protocol 240 through an Update() call issued by MRTE scheduler 245.

20 Upon updating MRTE protocol 240 with network status information, MRTE scheduler 245 makes an IndicateAdmit() call on MRTE service provider 210 to signal that admission to the network has been enabled. MRTE service provider 210 then forwards the IndicateAdmit() call to applications 205.

Figure 5 is a flowchart describing the general behavior of the QoS manager 225.
25 The behavior is implemented in software in one embodiment in the C++ language and comprises multiple objects which may or may not correspond precisely with the logical blocks used to describe the behavior. Further languages and different programming styles and hardware or firmware may be used in other embodiments as is well known in the art. The QoS manager is represented by start block or box 510; action boxes 515,
30 525, 535, 550, 555 and 565; and decision boxes 520, 530, 540, 545 and 560. The initialization of the QoS manager 225 is indicated by the arrival of a new traffic session in start box 510. The arriving session comprising a known number or frequency and size of packets to be sent to another node is scheduled within its QoS region in action box 515. A decision is made in decision box 520 as to whether the session is

schedulable. If the session is schedulable, flow is transferred to action box 550 for admission of the session.

If the session is not schedulable at this point, the QoS of other sessions may be reduced in action box 525. This process is highly dependent on the particular
5 application being serviced. There is usually a number of different levels of criticality of data associated with any application which are readily adaptable to QoS as is known in the art. An admission decision will then be made in decision box 530. If the session is schedulable, flow will be transferred to action box 550 for admission of the session. Further detail on the schedulability and admission decision processes is provided below.

10 If the session is not schedulable at this point, QoS manager 225 may suspend lower criticality sessions. An admission decision is then made in decision box 540. If the session is schedulable, flow is transferred to action box 550 for admission of the session.

If the session is not schedulable at this point, a decision must be made in
15 decision box 545 as to whether scheduling should be re-negotiated with a higher QoS for the session. If re-negotiation is required, flow is transferred to action box 515 to repeat the scheduling analysis. If re-negotiation is not required, flow is transferred to decision box 560 to be placed in a wait queue.

Upon admission of the traffic session at action box 550, the session will be sent
20 by action box 555. Once a session is sent by action box 555, decision box 560 will evaluate whether there are any unadmitted or waiting sessions. If there are waiting sessions, flow is transferred to action box 565 where the most critical waiting session is chosen. Flow is then transferred to action box 515 to repeat the scheduling analysis. If there are no waiting sessions, the process is concluded until new sessions arrive.

25 Figure 6 is a flowchart describing the general behavior of the MRTE scheduler 245. The MRTE scheduler is represented by start box 610; action boxes 615, 625, 630 and 635; and decision box 620. The initialization of the MRTE Scheduler 245 is indicated by a request for admission in start box 610. The request is analyzed according to deterministic scheduling algorithm 250 in action box 615. Upon determining the
30 schedulability in action box 615, MRTE scheduler 245 must decide, in decision box 620, whether the request can be guaranteed without adversely affecting already-admitted traffic.

If the request can be granted without adverse effects, MRTE scheduler 245 informs MRTE protocol 240 as to how the traffic is fragmented in action box 630.

MRTE scheduler 245 then grants the request in action box 635. If the request cannot be granted without adverse effects, MRTE scheduler 245 will deny the request at action box 625. Once the request is either granted or denied, the process terminates.

Figure 7 is a flowchart describing the general behavior of the MRTE protocol 240. The MRTE protocol is represented by start box 710; action boxes 715, 720, 730, 735, 750 and 755; and decision boxes 725, 740, 745 and 760. MRTE protocol 240 begins the process of network admission as indicated by an initialization in start box 710. Upon initialization of the MRTE protocol 240, system configuration information is gathered from the collision resolution protocol 235 as indicated in action box 715.

MRTE protocol 240 then waits for a user request and system interrupt in action box 720. A decision is made in decision box 725 as to whether one or both wait states have been satisfied. If both have been satisfied, flow is transferred to action box 730 to grant network occupation.

If both wait states of action box 720 have not been satisfied, MRTE protocol 240 decides if a user request was received in decision box 745. If not, flow will return to action box 720 to continue waiting. If a user request was received, MRTE protocol 240 then pre-processes the sending and admission request in action box 750. It then waits for the system interrupt in action box 755. Once the system interrupt is received, flow is transferred to action box 730 to grant network occupation.

Once network occupation is granted in action box 730 by either route, MRTE protocol 240 sends the data, processes the admission request and releases the network occupation in action box 735. It will then determine if other send requests are pending in decision box 740. If send requests are pending, flow is transferred to action box 750 for pre-processing. If no send requests are pending, flow is transferred to action box 720 to wait for further requests and system interrupts.

Figure 8 is a flowchart of the interrupt handling of MRTE protocol 240. MRTE protocol 240 is represented by start box 810; action boxes 830, 840, 850, 855, 860, 865, 875 and 880; and decision boxes 815, 820, 825, 835, 845 and 870. The process is initiated in start box 810 with receipt of an interrupt. MRTE protocol then decides in decision box 815 whether the interrupt is a token packet or TDMA interrupt from collision resolution protocol 235. If it is an interrupt from collision resolution protocol 235, MRTE protocol 240 decides whether an admission request is currently pending in decision box 820.

If the interrupt is not from collision resolution protocol 235, MRTE protocol 240 decides in decision box 845 whether the interrupt represents an MRTE private data packet containing MRTE system status information. If it is MRTE private data, MRTE protocol 240 updates its system status. If not MRTE private data, MRTE protocol 240
5 decides if it is a valid incoming MRTE data packet. If it is a valid data packet, MRTE protocol 240 passes the packet on the MRTE service provider 210 in action box 875. If the interrupt is not a valid MRTE data packet, MRTE protocol 240 passes the packet to the interested upper layer protocol, not defined by the invention.

If there are pending admission requests at decision box 820, MRTE protocol 240
10 will invoke MRTE scheduler 245 in action box 855, making the IndicatePermission() API call. After invoking MRTE scheduler 245, or if there are no pending admission requests, flow is transferred to decision box 825 to determine if there are any MRTE sending requests pending. If sending requests are pending, MRTE protocol sends the MRTE synchronous message packet in action box 860. After sending the synchronous
15 packet, or if there are no sending requests pending, flow is transferred to action box 830.

In action box 830, MRTE protocol 240 checks the timer and asynchronous message queue. Based on available time and the asynchronous message queue, MRTE protocol 240 decides in decision box 835 whether there is transmission time available and whether there are asynchronous messages ready. If time is available and
20 asynchronous data packets are ready for transmission, MRTE protocol 240 sends the asynchronous data packets in action box 865, then releases the Ethernet occupation in action box 840. If time is not available or there are no asynchronous data packets ready for transmission, MRTE protocol 240 simply releases the Ethernet occupation in action box 840.

Figure 9 is a flowchart of the admission control process of the MRTE scheduler 245 and comprises start box 910; action boxes 915, 925, 930, 935 and 940; and decision box 920. The process is initialized by the receipt of the IndicatePermission() call from MRTE protocol 240. Upon initialization, MRTE scheduler 245, in action box 915, obtains the network status information from the MRTE repository 255 using the Get()
25 call. MRTE scheduler 245 then determines if the send request can be scheduled by making the Schedule() call at 920 to the deterministic scheduling algorithm 250. If the request cannot be scheduled, MRTE scheduler 245 reports, in action box 940, to MRTE service provider 210 indicating admission failure using the IndicateReceive() call at
30 940. The scheduling and admission control algorithms are encapsulated in a

deterministic scheduling policy class, based on different protocols like token based or TDMA.

If deterministic scheduling algorithm 250 determines that the send request can be scheduled, MRTE scheduler 245 updates MRTE repository 255 in action box 925 using the Update() call. MRTE scheduler 245 then updates the MRTE protocol 240 with operation status information in action box 930 using the Update() call. Finally, MRTE scheduler 245 reports, in action box 935, to MRTE service provider 210 indicating admission success using the IndicateReceive() call.

There are two traffic models to be considered. The first is a periodic or synchronous message stream. The factors involved in this traffic model for each node 120 are its period (P_j); message length or transmission time (M_j); deadline (D_j); QoS (Q_j); and criticality or level of importance (C_j) where "j" represents an individual message stream. The second traffic model is an aperiodic or asynchronous message stream. The factors involved in this traffic model for each node 120 are the same as the periodic model with the elimination of the period (P_j).

In one embodiment, deterministic scheduling algorithm 250 utilizes a set of equations to determine if a request is schedulable in one embodiment. The relevant equations are as follows:

func {Eq.1: $TTRT = \min(P_j)/2$, 20

FORALL j}

func {Eq.2: $T_{NRT} = TTRT - T_{RT}$ }

func {Eq.3: $H_i = \sum_{j=1}^n \{m_j\}$

left lbrace $\{M_j\}$ over { left [$\{ \min(D_j, P_j) \}$ over {TTRT} right] ~ 1 } ~ + O_j right rbrace }

func { Eq.4: $\sum_{i=1}^n H_i + T_{NRT} \leq TTRT$

Where the following additional definitions apply:

TTRT: Target Token

Rotation Time

T_{RT} : Time interval for

transmitting real-time traffic

T_{NRT} : Time interval for

transmitting soft- or non-real-time traffic

30

I: Node number

j: Data stream number

H_i : Token holding time of individual node I

O_j : Software overhead of transmitting data stream j

n: Total number of nodes

m_i : Total number of real-time packets for transmission within H_i

A new request will be schedulable if Equation 4 is true, given Equations 1, 2 and

3.

5

CONCLUSION

A middleware approach to implementation of real-time Ethernet has been described which provides deterministic, i.e. predictable, communication services for both real time and non-real time traffic over a conventional Ethernet network having a plurality of nodes desiring to transmit packets of data. The middleware comprises computer software residing above a network interface device and the device driver, yet below the system transport services and/or user applications. The invention provides a Middleware Real-Time Ethernet or MRTE which does not require modification to existing hardware that implements Ethernet. Separate queues are used for deterministic scheduling to determine the order of packet queuing and transmission on each node such that (1) real-time traffic can be guaranteed once admitted for transmission service, (2) non-real-time traffic can be served, and (3) the Ethernet bandwidth utilization can be optimized. Quality of Service, QoS, enables making on-line tradeoffs between network bandwidth availability and network transmission quality. Examples of QoS include (1) degree of packet collisions when Ethernet is shared by soft- or non-real-time traffics during certain time slots and (2) amount of end-to-end packet transmission latency.

When QoS is used, periodic data, such as video at 30 frames per second may be given a priority or criticality, and a cumulative loss factor, e.g. up to four frames in a row may be discarded. If there is sufficient bandwidth remaining after higher priority tasks or data streams are handled, the video will be accepted to the real time queue with at least five frames per second being sent. If other tasks are deleted or reduced, this frame rate will increase.

Software structuring enables hosting of the real-time Ethernet middleware above the Ethernet network device and the device driver, and below system transport software and/or user applications. A specific example of such a software host is the Microsoft® Network Device Interface Specification (NDIS) with Device Driver Kit (DDK) on Microsoft® NT®-based personal computer platforms. Many other software hosts are available depending upon specific hardware chosen.

The real-time Ethernet middleware comprises two main function modules, a collision avoidance module and a deterministic scheduling module. The collision

avoidance module implements a collision-avoidance protocol that provides the capability for preventing Ethernet traffic from colliding, which is one source of the problem of non-deterministic Ethernet behavior. A specific example of such a protocol is a token-based protocol by which a token circulating among the Ethernet nodes
5 determines which node should transmit packets at any point in time. Other collision-avoidance protocols may be used with the invention such as various implementations of Time-Division Multiple Access (TDMA), a technology using Time-Division Multiplexing (TDM). The protocol or standard must merely provide a mechanism to avoid conflict among data transmissions by more than one node at any given time.
10 Either embodiment provides benefits for real time process control, multimedia and Internet applications as well as other applications which might depend on arrival of real time traffic.

The deterministic scheduling module determines if a set of real-time traffic in the entire distributed system can be guaranteed with respect to their timing constraints,
15 such as end-to-end transmission latency.

In one embodiment, the collision-avoidance protocol is switchable to be enabled or disabled as desired by the deterministic scheduling module. This allows the invention to guarantee no collisions of real-time traffic, yet permit collisions of soft- and non-real-time traffic. Such mixed-mode operation could lead to increased bandwidth
20 utilization depending upon the loading during time periods allocated to soft- and non-real-time traffic. Lightly loaded CSMA/CD systems can be more efficient than systems operating on a collision avoidance protocol.

While the Collision-Avoidance Protocol is active, the time set for a complete rotation of transmitting nodes is bounded. In the case of a token-based protocol, the
25 token must return within this bounded time, or token rotation time.

In a further embodiment of the invention, allocation of bandwidth to an individual bridge or node is increased based on underutilization of bandwidth by other bridges or nodes in the network.

While the invention was described in connection with various embodiments, it
30 was not the intent to limit the invention to one such embodiment. Many other embodiments will be apparent to those of skill in the art upon reviewing the above description. In one embodiment of the invention, the QoS module is eliminated. Due to the modular nature of the invention, the system is capable of accepting a QoS module at some later time if desired by the user.

In another embodiment of the invention, multiple local Ethernet networks may be bridged together. Each bridge between networks would accept and schedule messages or streams of data. The data streams held by an individual bridge would be sent when that bridge is designated to transmit, such as when the bridge is in possession of the token in a token-based protocol. New data streams may be refused based on whether the bridge would have sufficient bandwidth to send the data after sending all higher priority messages. To guarantee no collisions for a given period of time, all bridges must be operating in the mode of the collision-avoidance protocol during that period.

What is claimed is:

1. A method of communicating real time traffic on a collision detection based communication network comprising the steps of:
5 queuing real time traffic separate from non-real time traffic;
 sending real time traffic during a first time interval of a communication cycle while using a deterministic scheduling protocol; and
 sending non-real time traffic during a second time interval of the communication cycle.
10
2. A machine readable medium having instructions stored thereon for causing a computer to implement the steps of claim 1.
3. A method of communicating real time traffic on a collision detection based communication network comprising the steps of:
15 receiving a request at a node coupled to the communication network indicating that it has real time traffic to send;
 determining if a given quality of service can be provided for the real time traffic;
 adjusting a first amount of time per communication cycle granted to non-real
20 time traffic if said quality of service cannot be provided; and
 accepting the real time traffic request.
4. A machine readable medium having instructions stored thereon for causing a computer to implement the steps of claim 3.
25
5. The method of claim 3 and further comprising the steps of:
 queuing the real time traffic separate from non-real time traffic;
 sending real time traffic during the first amount of time of the communication cycle while using a deterministic scheduling protocol; and
30 sending non-real time traffic during a second amount of time of the communication cycle.
6. A machine readable medium having instructions stored thereon for causing a computer to implement the steps of claim 5.

7. A node coupled to a collision detection based communication network comprising:

a first queue that queues real time traffic;

5 a second queue that queues non-real time traffic;

a scheduler that schedules real time traffic for sending over the communication network during a first period of time per communication cycle using a deterministic protocol and that provides non-real time traffic for sending over the communication network during a second period of time per communication cycle.

10

8. The node of claim 7 and further comprising a quality of service manager that modifies the first and second periods of time based on the type of traffic generated by the node.

15

9. The node of claim 7 wherein the collision detection based communication network comprises Ethernet, and the collision avoidance protocol comprises a token ring or time division protocol.

10. A communication network comprising:

20

a plurality of nodes coupled by Ethernet implementing hardware, each node comprising:

a middleware set of software modules that serve real time traffic during a first period of a communication cycle using a deterministic protocol to the hardware, and non-real time traffic during a second period of the communication cycle to the hardware for normal Ethernet transmission such that all the nodes operate using the same protocols during each period.

25

11. A communication protocol for an Ethernet network for transmission of real-time and non-real-time data packets, the Ethernet network containing network devices, device drivers, system network transport software and user applications, the communication protocol comprising:

30

a software structuring module for hosting the communication protocol above the network devices and device drivers, and below system network transport software or user applications;

a deterministic scheduling module for determining the schedulability and order of admission of data packets for transmission across the Ethernet network; and

a collision-avoidance protocol module for preventing collision among Ethernet data packets as required by the deterministic scheduling module.

5

12. The communication protocol of claim 11, wherein the collision-avoidance protocol module is token based.

10

13. The communication protocol of claim 11, wherein the collision-avoidance protocol module is time-division multiple access based.

14. The communication protocol of claim 11, wherein the deterministic scheduling module prohibits collision of real-time data packets and permits collision of non-real-time data packets.

15

15. The communication protocol of claim 11, further comprising:
a quality of service module for making on-line tradeoffs between the Ethernet network availability and the Ethernet network transmission quality.

20

16. The communication protocol of claim 11, wherein the deterministic scheduling module determines schedulability of real-time data packets according to the following equations:

func {Eq.~1:~TTRT~min(P_j)/2,~	
FORALL j}	25 where:
func {Eq.~2:~T_NRT~TTRT~T_RT}	TTRT is the target token
func {Eq.~3:~H_i~sum from {j=1} to	rotation time;
{m_i}~	T_RT is the time interval for
left lbrace ``{M_j} over { left [{min`` (D_j ,	transmitting real-time traffic;
``P_j``} over {TTRT} right]~1}	30 T_NRT is the time interval for
~+~ O_j`` right rbrace}	transmitting soft- or non-real-time
func { Eq.~4:~ sum from {i=1} to	traffic;
n``H_i~+~T_{NRT}~TTRT}	I is the node number;
j is the data stream number;	

H_i is the token holding time of individual node I ;

O_j is the software overhead of transmitting data stream j ;

n is the total number of nodes; and

m_i is the total number of real-time packets for transmission within H_i ; and

5 wherein a real-time data packet is schedulable if Equation 4 evaluates to true given Equations 1, 2 and 3.

17. A communication protocol for an Ethernet network for transmission of real-time and non-real-time data packets comprising a deterministic scheduling algorithm wherein
10 the deterministic scheduling algorithm:

will guarantee transmission of both real-time and non-real-time data packets;

will permit transmission of real-time data packets only if such packet will not conflict with transmission of other data packets; and

will optimize utilization of the Ethernet network.

15

18. The communication protocol of claim 17, wherein the deterministic scheduling algorithm comprises the following equations:

func {Eq.~1:~ $T_{TRT} = \min(P_j)/2$,~

FORALL j } 20

func {Eq.~2:~ $T_{NRT} = T_{TRT} - T_{RT}$ } where:

func {Eq.~3:~ $H_i = \sum \text{from } \{j=1\} \text{ to } \{m_i\}$ } T_{TRT} is the target token
rotation time;

left lbrace `` $\{M_j\}$ over { left [\min `` $\{D_j$, T_{RT} is the time interval for
`` P_j `` } over { T_{TRT} } right] ~ ~ 1 } 25 transmitting real-time traffic;

~ ~ O_j `` right rbrace} T_{NRT} is the time interval for

func { Eq.~4:~ $\sum \text{from } \{i=1\} \text{ to } n$ `` $H_i + T_{NRT}$ `` } T_{TRT} transmitting soft- or non-real-time
traffic;

I is the node number;

30 j is the data stream number;

H_i is the token holding time of individual node I ;

O_j is the software overhead of transmitting data stream j ;

n is the total number of nodes; and

m_i is the total number of real-time packets for transmission within H_i ; and

such that a real-time data packets will be permitted transmission only if Equation 4 evaluates to true given Equations 1, 2 and 3.

5 19. A communication protocol for an Ethernet network for transmission of real-time and non-real-time data packets, the Ethernet network containing network devices, device drivers, system transport software and user applications, the communication protocol comprising:

10 software structuring means for hosting the communication protocol above the network devices and device drivers, and below system transport software and user applications;

deterministic scheduling means for scheduling admission of data packets for transmission across the Ethernet network; and

collision-avoidance protocol means for preventing collision among Ethernet data packets as required by the deterministic scheduling means.

15

20. The communication protocol of claim 19, wherein the deterministic scheduling means prohibits collision of real-time data packets and permits collision of non-real-time data packets.

20

21. The communication protocol of claim 19, further comprising:

quality of service means for making on-line tradeoffs between the Ethernet network availability and the Ethernet network transmission quality.

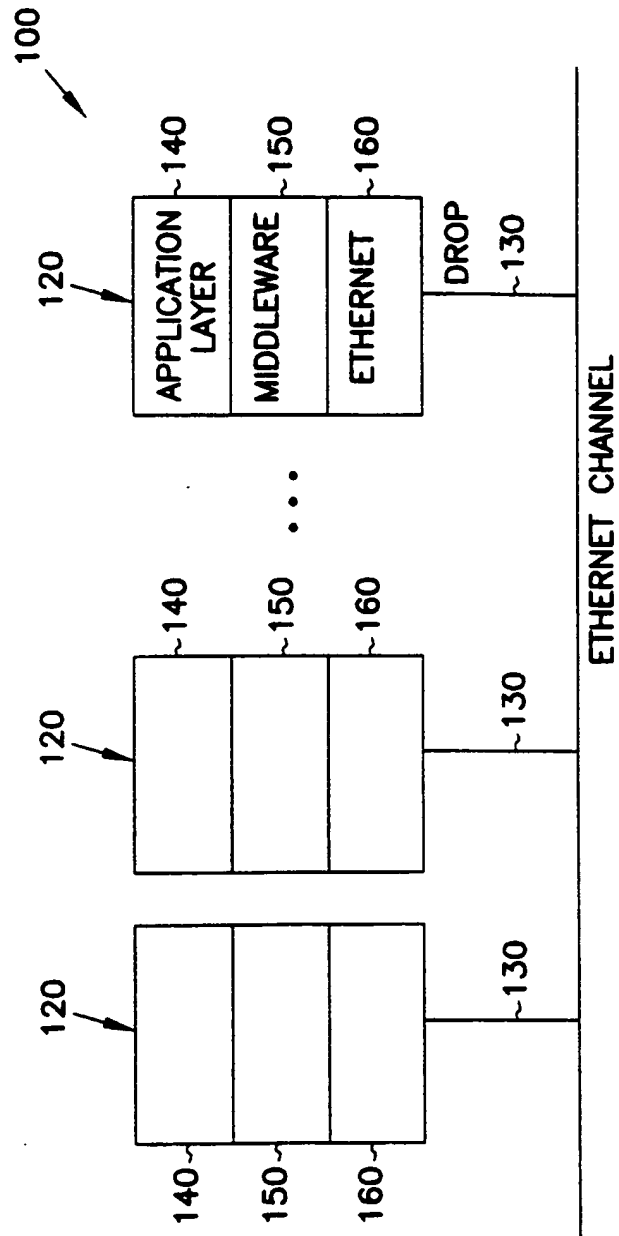


FIG. 1

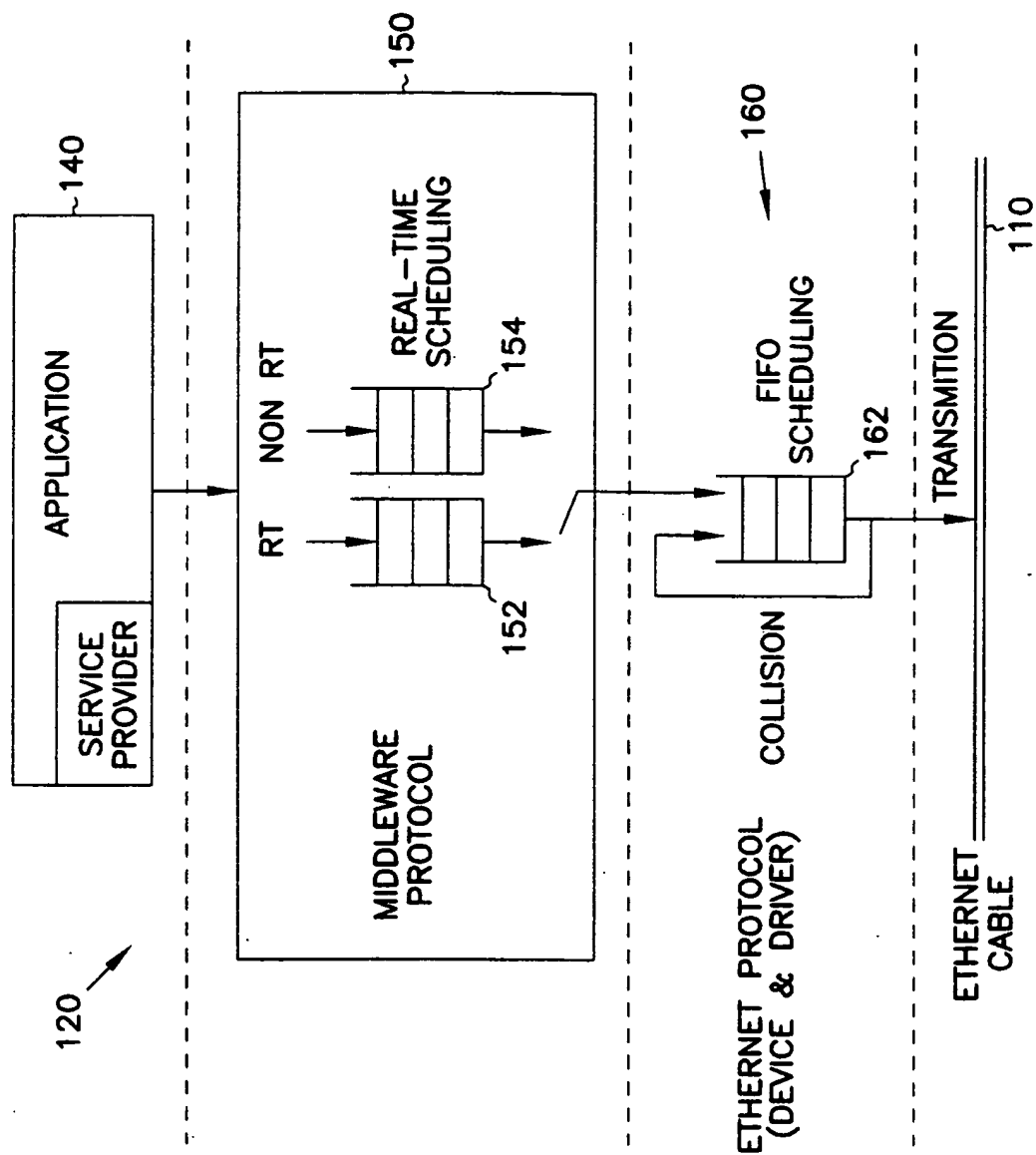


FIG. 2A

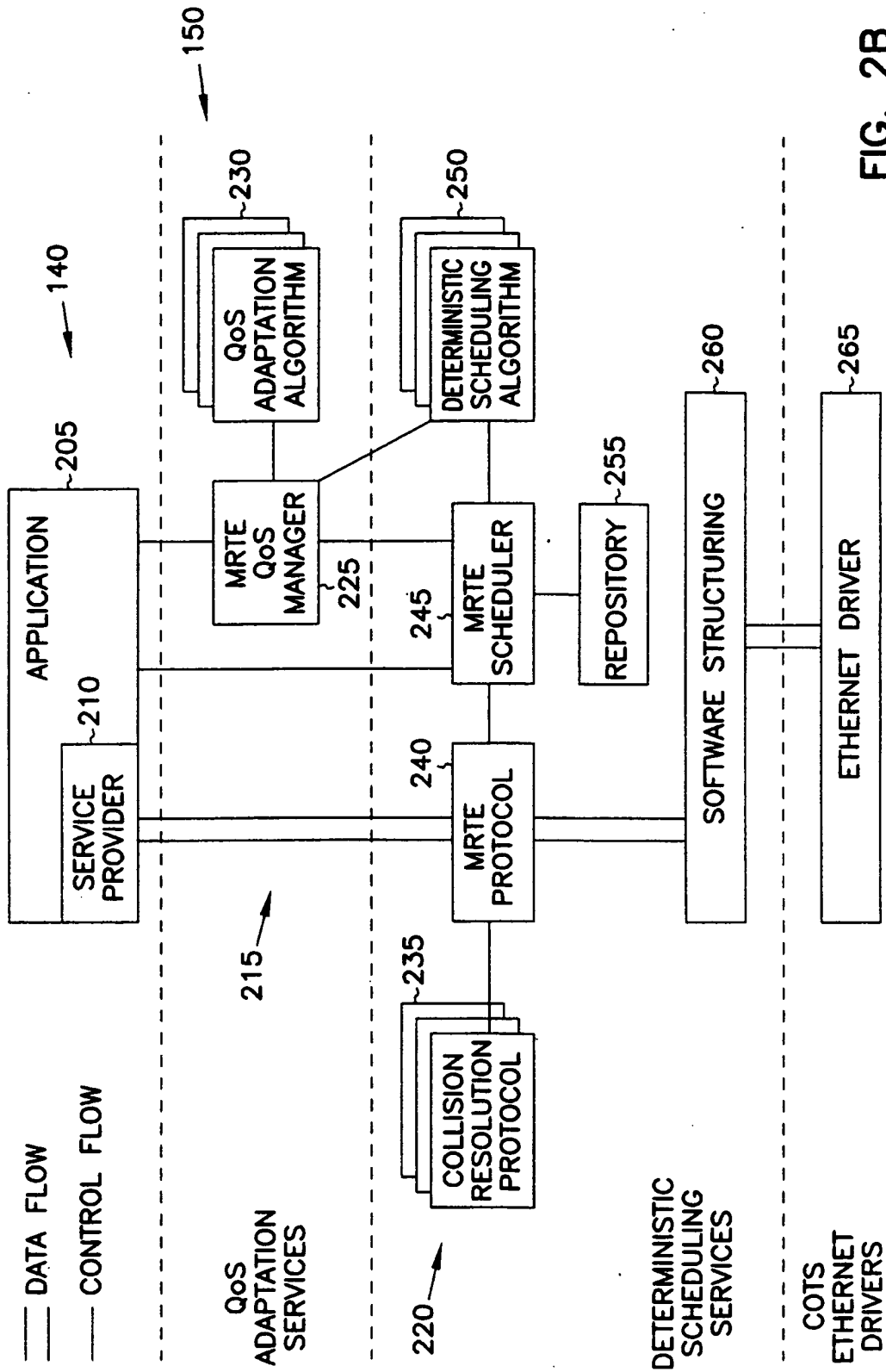


FIG. 2B

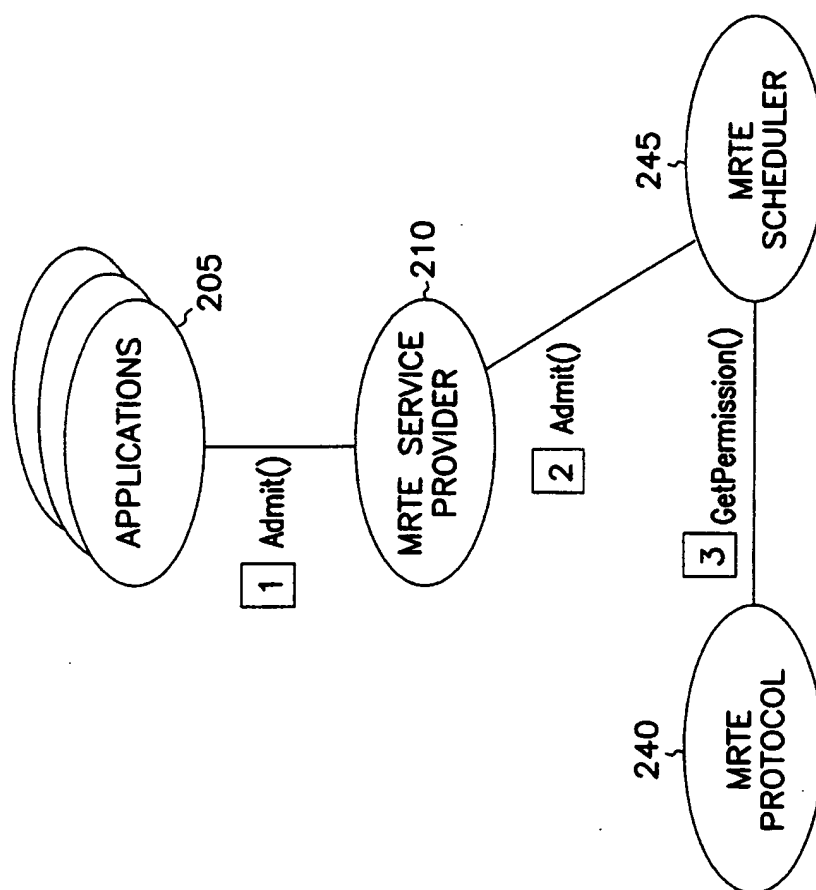
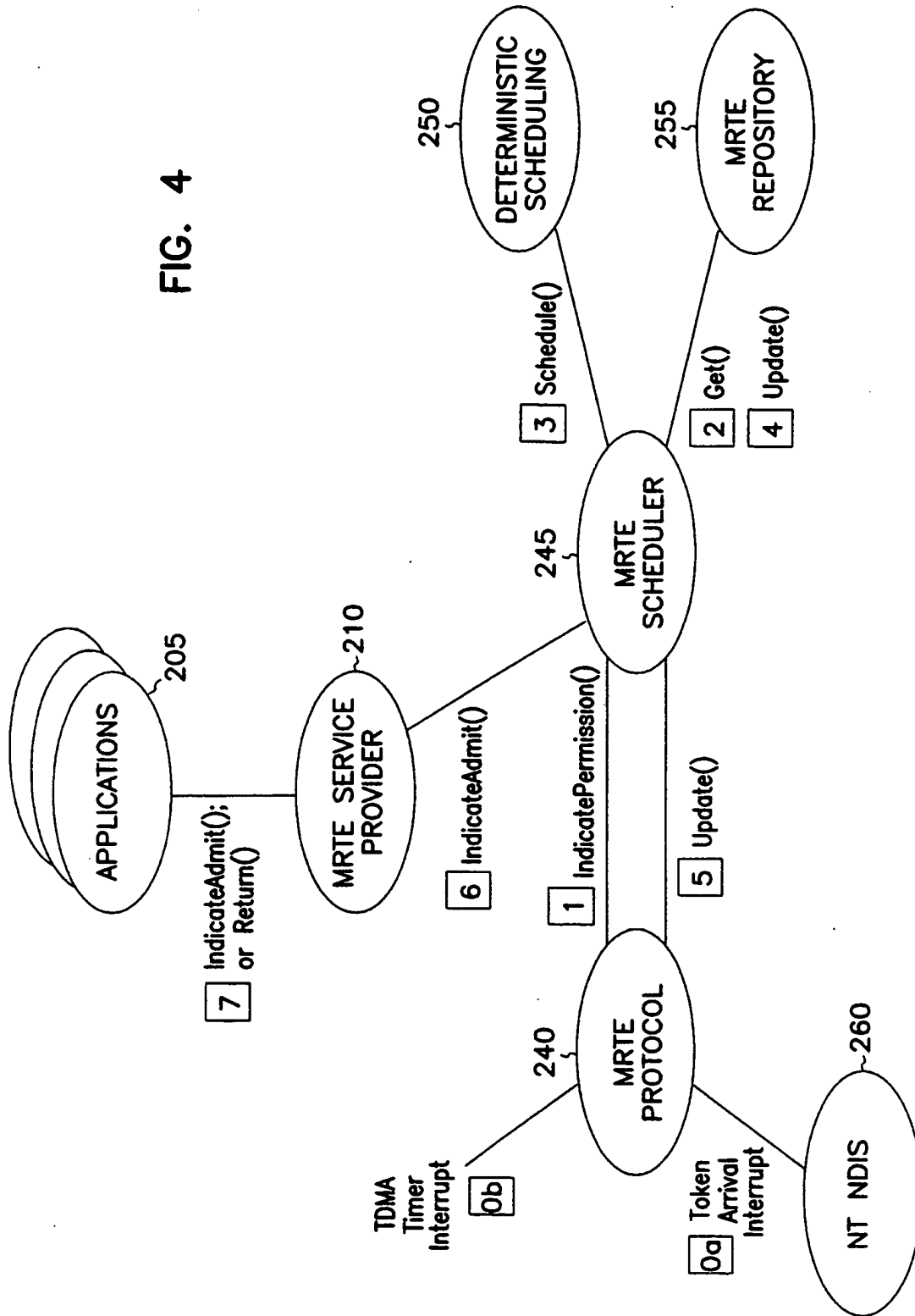


FIG. 3

FIG. 4



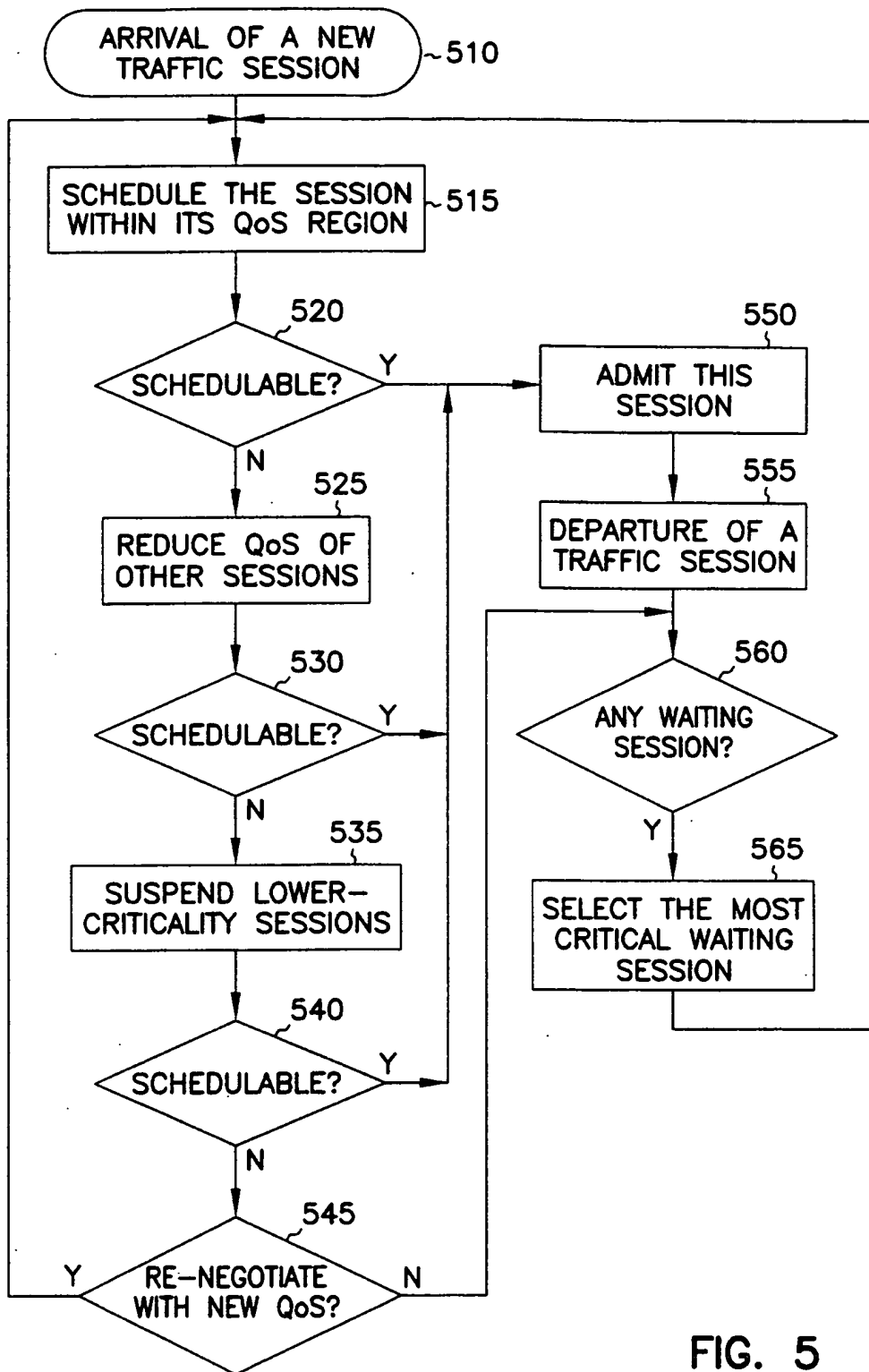


FIG. 5

FIG. 6

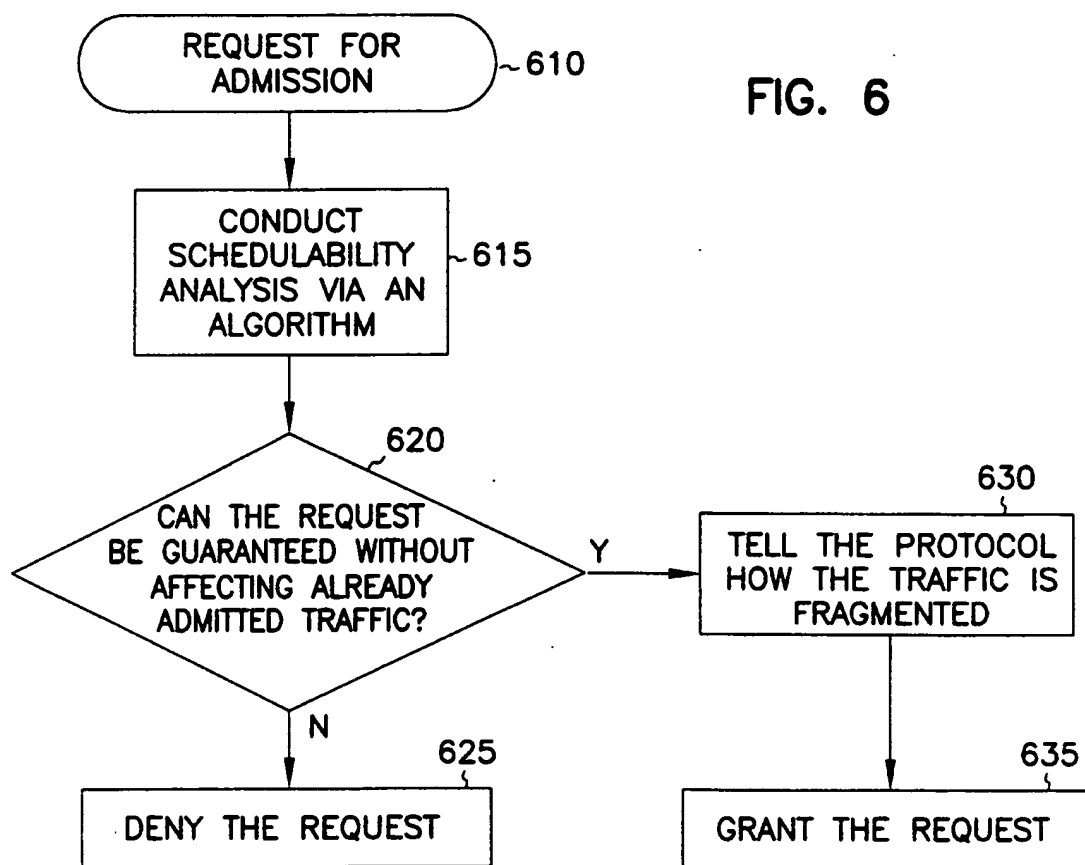
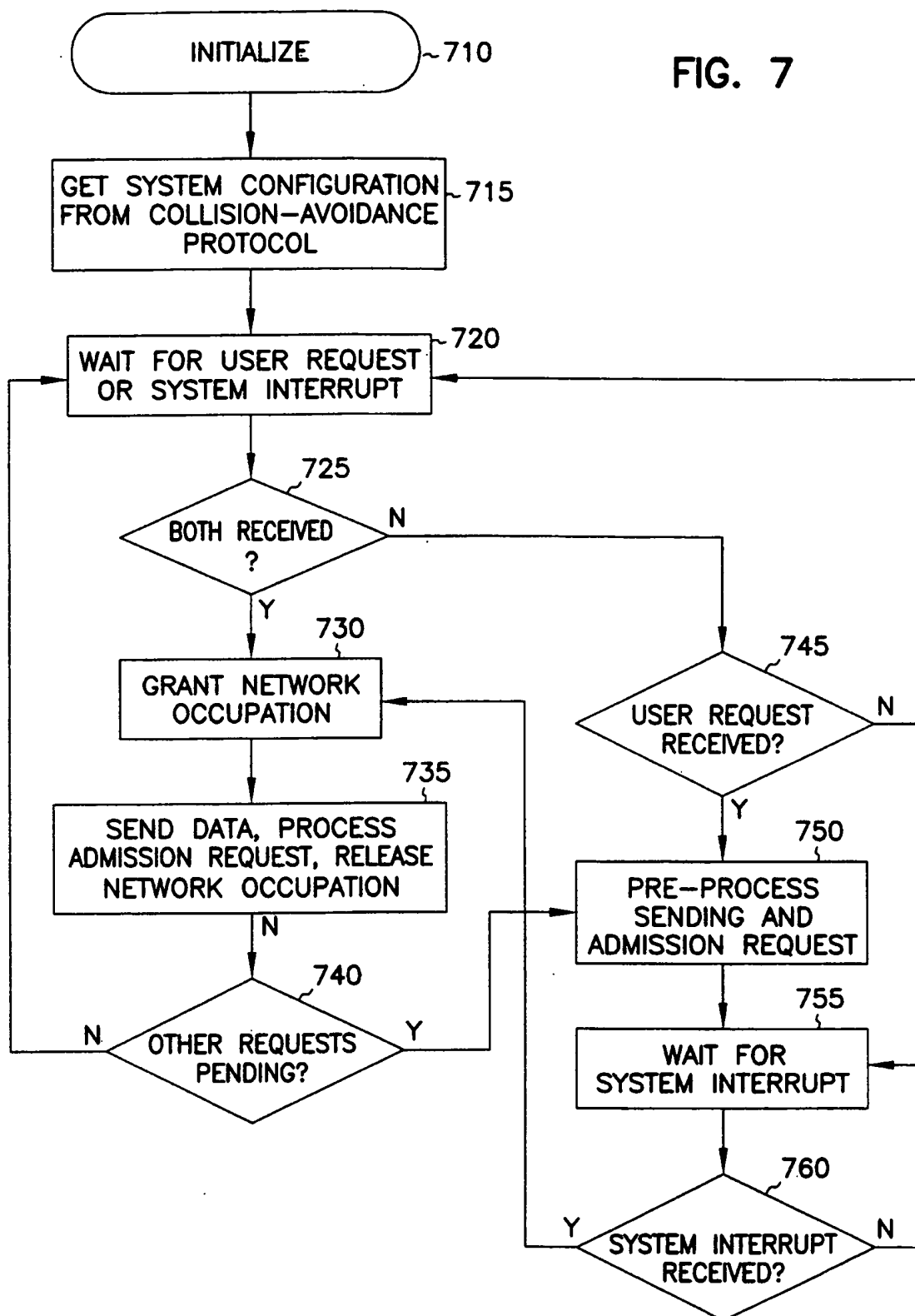


FIG. 7



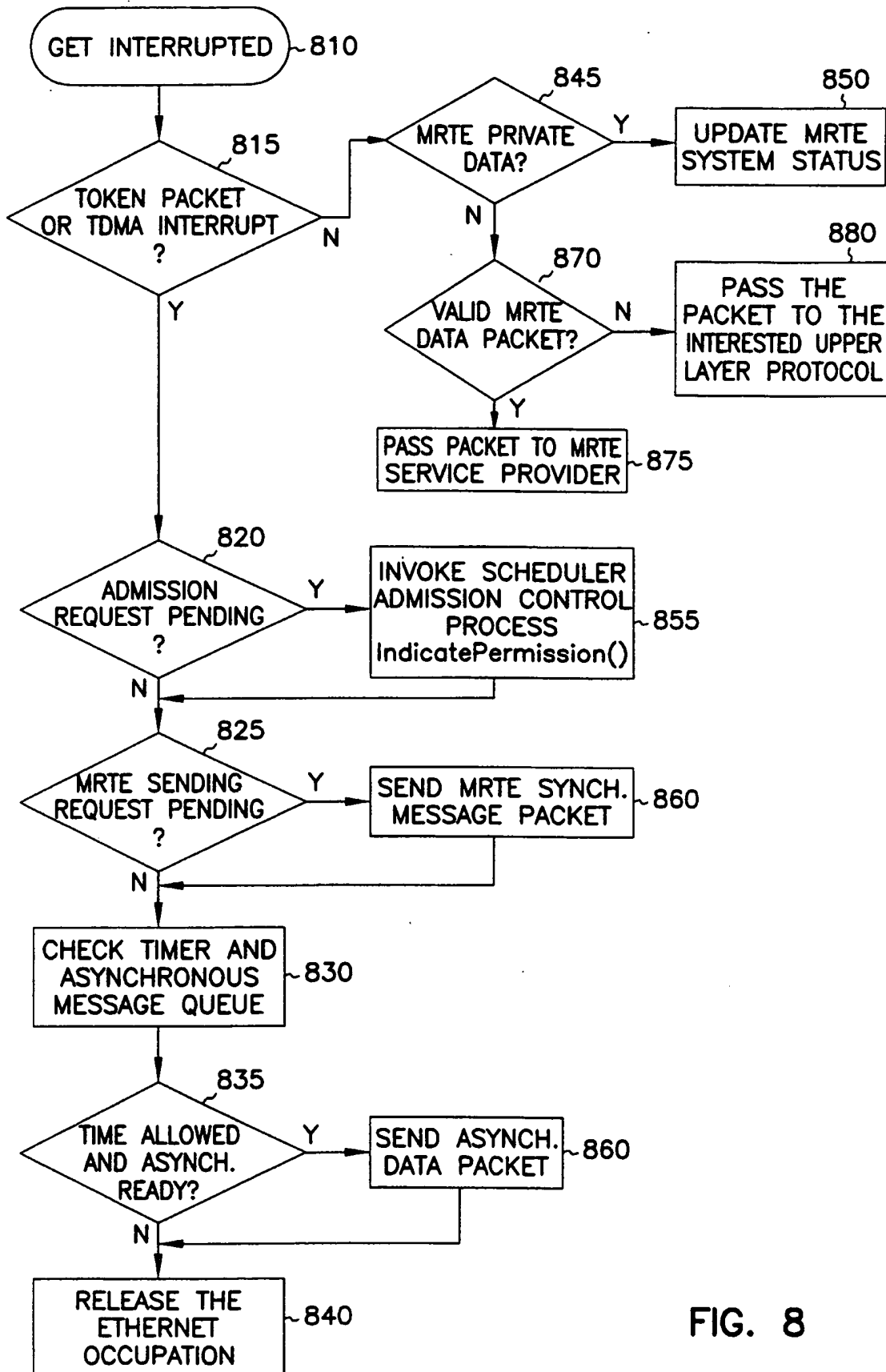


FIG. 8

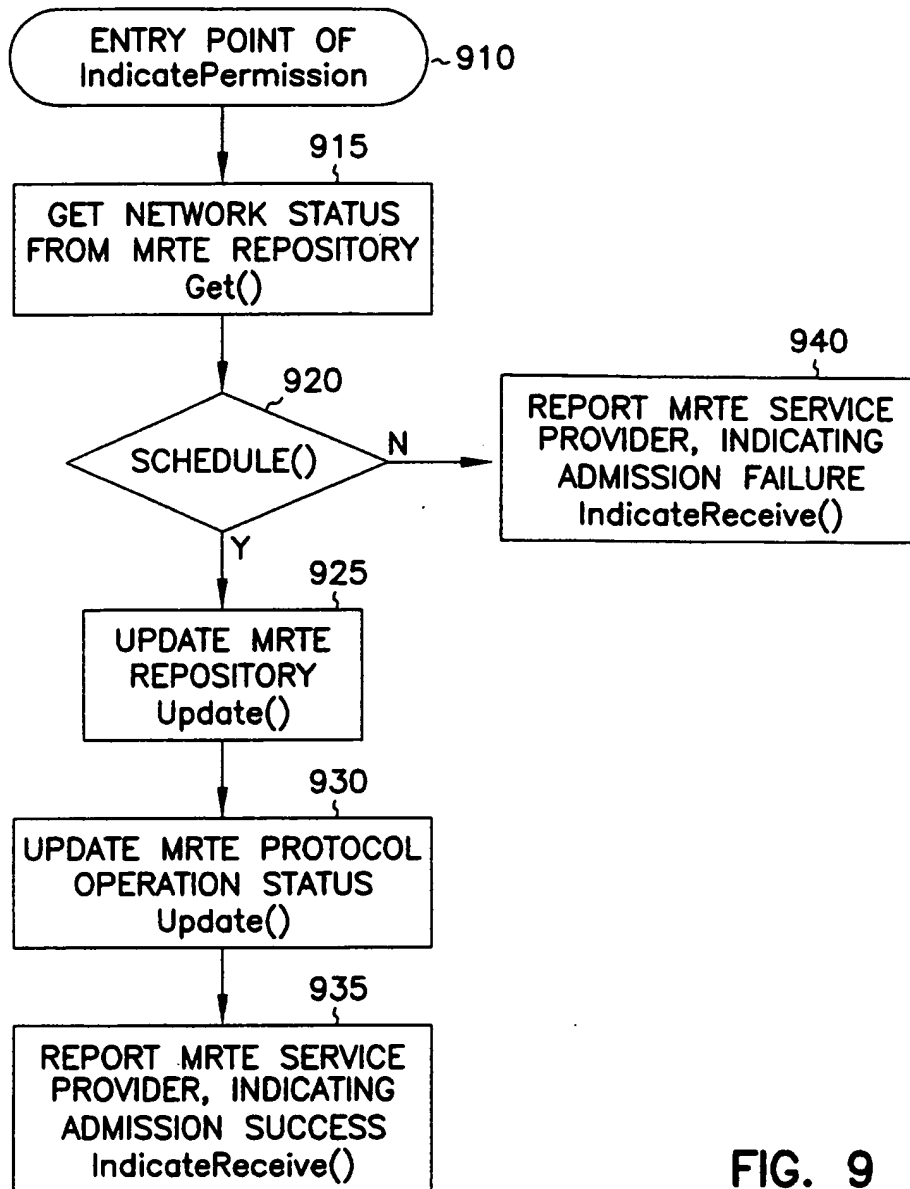


FIG. 9

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/14083

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04L12/64 H04L12/413

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 761 430 A (LIEB DEREK WEARIN ET AL) 2 June 1998 (1998-06-02)	1-11, 13-15, 17,19-21
Y	the whole document	12
A		16,18
X	SHARROCK S M ET AL: "A CSMA/CD-BASED, INTEGRATED VOICE/DATA PROTOCOL WITH DYNAMIC CHANNEL ALLOCATION" COMPUTER NETWORKS AND ISDN SYSTEMS, vol. 18, no. 1, 24 November 1989 (1989-11-24), pages 1-18, XP000070488 ISSN: 0169-7552	1-9,17
A	page 3, right-hand column, paragraph 2 -page 6, left-hand column, paragraph 2 page 14, right-hand column, paragraph 3 -page 16, right-hand column, paragraph 2 --- -/--	10-16, 18-21

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

20 October 1999

Date of mailing of the international search report

05/11/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

RAMIREZ DE AREL..., F

INTERNATIONAL SEARCH REPORT

Intern: al Application No
PCT/US 99/14083

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHITRA VENKATRAMANI ET AL: "DESIGN, IMPLEMENTATION, AND EVALUATION OF A SOFTWARE-BASED REAL -TIME ETHERNET PROTOCOL" COMPUTER COMMUNICATIONS REVIEW, vol. 25, no. 4, 1 October 1995 (1995-10-01), pages 27-37, XP000541648 ISSN: 0146-4833	12
A	page 28, right-hand column, paragraph 3 -page 32, left-hand column, paragraph 1 ---	10,11,19
A	SZABO C: "AN ETHERNET COMPATIBLE PROTOCOL TO SUPPORT REAL TIME TRAFFIC AND MULTIMEDIA APPLICATIONS" COMPUTER NETWORKS AND ISDN SYSTEMS, vol. 29, no. 3, 1 February 1997 (1997-02-01), pages 335-342, XP000642427 ISSN: 0169-7552 page 336, left-hand column, paragraph 2 -page 338, left-hand column, paragraph 1 -----	1-21

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat l Application No

PCT/US 99/14083

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5761430 A	02-06-1998	AU 2453397 A	07-11-1997
		EP 0832530 A	01-04-1998
		WO 9739556 A	23-10-1997
<hr/>			